

Bases de Données

Examen (extraits)*

Luc Segoufin, Emmanuel Waller

16 Décembre 2003

Durée de l'examen: 2 heures. Les transparents du cours et les énoncé de TD sont autorisés. Tout autre document est interdit. Écrivez lisiblement. Justifiez vos résultats.

1 Applications du cours

1.1 Expressivité

L'opérateur division noté "div" est ajouté à l'algèbre relationnelle. Pour toutes instances I, J avec $sort(J) \subseteq sort(I)$, $I \div J$ est une instance sur $sort(I) - sort(J)$ définie par:

$$I \div J = \{u \mid (\{u\} \times J) \subseteq I\}.$$

1. Donner une requête dans l'algèbre relationnelle étendue avec la division calculant le nom des viticulteurs vendant tous les vins de Bourgogne.
2. Exprimer la même requête sans utiliser la division.
3. Est-ce que la division est une opération monotone?
4. Prouver que toute requête dans l'algèbre relationnelle étendue avec la division peut être exprimée dans l'algèbre relationnelle (sans la division).

1.2 Équivalences

Montrer si chacune des propositions suivantes est vrai ou fausse:

1. $\pi_X(I \cup J) = \pi_X(I) \cup \pi_X(J)$
2. $\pi_X(I \cap J) = \pi_X(I) \cap \pi_X(J)$
3. $\pi_X(I - J) = \pi_X(I) - \pi_X(J)$
4. $\pi_{AB}(\sigma_{B=5}(I)) \bowtie \pi_{BC}(\pi_{AB}(I) \bowtie \pi_{AC}(\sigma_{B=5}(I))) = \pi_{AB}(\sigma_{B=5}(I)) \bowtie \pi_{BC}(\sigma_{B=5}(I)).$

*Voyez le texte complet sur la page du cours

5. On considère les q_i donnés par :

$$\begin{aligned} q_0 : res(x, y) &\leftarrow R(x, y) \\ q_1 : res(x, y) &\leftarrow R(x, y_1), R(x_1, y_1), R(x_1, y) \\ q_2 : res(x, y) &\leftarrow R(x, y_1), R(x_1, y_1), R(x_1, y_2), R(x_2, y_2), R(x_2, y) \\ q_3 : res(x, y) &\leftarrow R(x, y_1), R(x_1, y) \end{aligned}$$

Est-ce que $q_i \subseteq q_j$ pour i, j in $[0..4]$ (justifier)?

2 Problèmes : dépendances fonctionnelles

Le problème porte sur les dépendances fonctionnelles qui jouent un rôle crucial dans l'aide à la conception de schéma de base de données. Les étoiles indiquent la difficulté de certaines parties. Les questions étoilées peuvent toutes être sautées dans un premier temps. Elles sont plus dures que les autres et rapporteront donc beaucoup de points.

On considère une relation I sur un ensemble d'attributs U . Une *dépendance fonctionnelle* sur U est une expression de la forme $X \rightarrow Y$ où X et Y sont des sous-ensembles de U . Une relation I sur U *satisfait* la dépendance $X \rightarrow Y$ si quels que soient les n -uplets u et v de I ,

$$\pi_X(u) = \pi_X(v) \text{ implique } \pi_Y(u) = \pi_Y(v).$$

(On dit aussi que X *détermine* Y .)

Soient $U_1 = ABCDE$, $F_1 = \{A \rightarrow C, B \rightarrow C, CD \rightarrow E\}$ et

	A	B	C	D	E
I_1	1	2	3	4	7
=	3	2	3	4	7
	2	2	3	4	7
	4	2	3	4	7
	4	2	3	8	8

1. Montrer que I_1 ne satisfait pas les dépendances fonctionnelles $AB \rightarrow D$ et $E \rightarrow A$. Est-ce que I_1 satisfait les dépendances de F_1 ?

Soient $X, Y \subseteq U$ et F un ensemble de dépendances fonctionnelles sur U , on dit que F *implique* $X \rightarrow Y$ si la propriété suivante est vérifiée :

pour chaque relation I sur U , si I satisfait F alors I satisfait aussi $X \rightarrow Y$.

La *fermeture* de X pour F notée $(X, F)^*$ est l'ensemble des attributs A de U tels que F implique $X \rightarrow A$.

2. Calculer $(AB, F_1)^*$, $(CD, F_1)^*$ et $(CDE, F_1)^*$.
3. Montrer que F implique $X \rightarrow Y$ si et seulement si $Y \subseteq (X, F)^*$.
4. Montrer que si $Y \rightarrow Z$ est dans F et $Y \subseteq (X, F)^*$, alors $Z \subseteq (X, F)^*$.

5. Supposant que I satisfait F_1 , donner un plan d'évaluation performant de la requête

$$\sigma_{D=5}(\pi_{ADE}(U) \bowtie (\sigma_{A=2}(\pi_{ABD}(U) \bowtie \pi_{AC}(U))))).$$

Quelle structure d'accès doit-on avoir pour que l'on puisse répondre à cette requête en temps constant ?